



Clewer, DR., Luo, L., Canagarajah, CN., Bull, DR., & Barton, MH. (2001). Efficient multiview image compression using quadtree disparity estimation. In *IEEE International Symposium on Circuits and Systems (ISCAS 2001) Sydney, NSW, Australia* (Vol. 5, pp. 295 - 298). Institute of Electrical and Electronics Engineers (IEEE).  
<https://doi.org/10.1109/ISCAS.2001.922043>

Peer reviewed version

Link to published version (if available):  
[10.1109/ISCAS.2001.922043](https://doi.org/10.1109/ISCAS.2001.922043)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# EFFICIENT MULTIVIEW IMAGE COMPRESSION USING QUADTREE DISPARITY ESTIMATION

*D. R. Clewer, \*L. J. Luo, C. N. Canagarajah, D.R. Bull, M.H. Barton*

Centre for Communications Research, University of Bristol,  
Woodland Road, Bristol BS8 1UB, UK.

## ABSTRACT

Viewpoint adaptation, or 'look-around' capability, is likely to be vital to the success of future immersive multimedia services such as 3D video-conferencing, virtual reality and 3DTV. This paper proposes a low complexity and efficient compression system for multiview images using quadtree disparity estimation and multiview synthesis. Previous approaches to the problem have generally involved the transmission of a coded stereo image pair, usually using fixed block size disparity compensation, along with a pair of dense disparity maps used for intermediate view synthesis. In this work, the quadtree disparity map coded implicitly in the stereo compression process is re-used in the multiview synthesis process. The other map is also efficiently represented using a quadtree structure at low resolution. Improved multiview compression results are presented, along with intermediate views synthesised at the decoder.

## 1 INTRODUCTION

Future advances in interactive television, mobile and Internet communications will not only enable a higher quality of service, but will enable new immersive and interactive multimedia services to be supported. One such immersive application is stereoscopic and multiview images and video. Indeed, high quality stereoscopic viewing equipment is already commercially available, and autostereoscopic displays (without viewing aids) are under development.

Previous approaches to the problem of multiview image coding have generally involved the transmission of a stereo image pair, along with a pair of dense disparity maps describing the stereo correspondence between the two images [1,2]. At the decoder these maps are used to synthesise intermediate views between the original images, thus supporting a look-around effect when viewed using appropriate equipment. The tasks of stereo image compression and dense disparity estimation have generally been kept separate.

This paper proposes a technique where the tasks of stereo image compression and the creation of dense disparity maps are partly combined using quadtree disparity estimation, leading to increased compression efficiency and much reduced computational complexity. The paper is organised as follows.

The work described in section 2 attempts to improve the efficiency of stereo image compression by using variable block size disparity estimation, thus reducing the prediction error, whilst keeping the overhead of the increased disparity and segmentation information to a minimum. The quadtree structure, as has been previously suggested [3], is found to be an efficient way of representing the segmentation information.

Furthermore, the increased resolution and accuracy of the disparity field around object boundaries allows for the possibility of intermediate view synthesis, a technique normally requiring dense disparity maps produced independently of the compression process. Section 3 demonstrates the results of intermediate view synthesis achieved using disparity maps produced using the quadtree-based disparity estimation technique. A prediction configuration is used where the right-to-left disparity map is transmitted as part of the compression process, and a low resolution left-to-right map as a small overhead. Finally, in section 4, the conclusions are presented, and areas for further investigation are suggested.

## 2 STEREO IMAGE CODING USING QUADTREE DISPARITY ESTIMATION

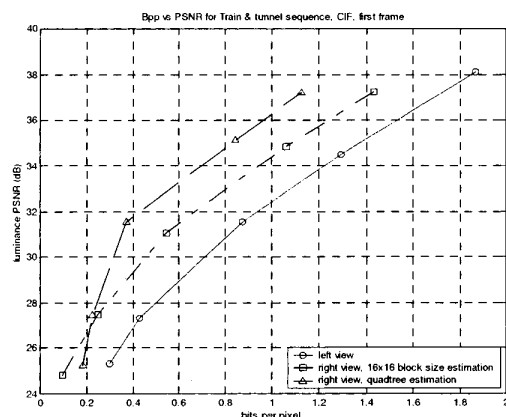
Previous approaches to the problem of stereo compression have often used fixed block size disparity estimation and compensation [4,5], similar to the motion compensation technique used for video coding. However, it has been shown [6] that, for low and medium activity scenes, block based disparity estimation is generally less efficient than motion estimation for the same block size. This is due in part to the change in perspective between the two viewpoints, causing uncovered background to arise at the boundaries of objects across the whole image plane. In

---

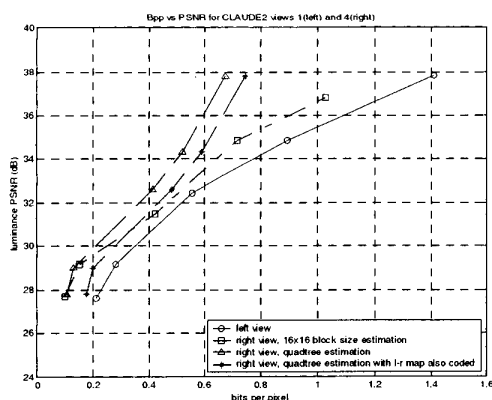
\* Dr. L.J.Luo is now with Oxford Semiconductor Ltd, Abingdon, Oxfordshire, UK.

comparison, with motion estimation there is often only a few localised objects in motion against a static background, so uncovered background occurs less frequently.

By splitting large block sizes into smaller ones around the boundaries of objects and in other areas of high prediction error, one would expect the overall prediction error energy to be reduced. Some overhead is required, however, to convey this extra segmentation information to the receiver. The quadtree structure is a well known and efficient way of achieving this, and has been used previously in image and video coding research [7]. Quadtree disparity estimation has also been previously suggested for improved stereo compression [3], but results have only been presented at very low bit rates and peak signal-to-noise (PSNR) values. In this work, experiments were carried out on several images at a wide range of bit rates.



(a) Train & Tunnel



(b) Claude2

Figure 1 Comparison of stereo image compression performance: quadtree disparity estimation vs. fixed block size estimation.

Figure 1 shows a comparison of the performance of stereo image compression using quadtree disparity estimation against fixed block size disparity estimation. The images were all of size 352x288 pixels, using the YUV 4:2:0 colour format. The fixed block size was 16x16 pixels, and for the quadtree estimation, a maximum of 16x16 and a minimum of 2x2 was used. Blocks were split if the minimum mean squared prediction error exceeded a pre-determined threshold. This threshold was manually adjusted to achieve different points on the rate-distortion curve. For the quadtree technique to be widely used, some automatic rate control methods would need to be developed.

The left image was intra-coded using standard DCT, quantization and entropy coding. The prediction error was also coded in this manner. The disparity vectors and overhead information was differentially and Huffman coded using look up tables adapted from the MPEG-2 standard. The results clearly show the improved compression performance of the quadtree technique over fixed block sizes for the images used in this work.

Table 1 shows a comparison of the number of bits spent on coding the stereo image pair for the fixed block size and quadtree methods for the Train & Tunnel image pair.

Method	R PSNR (dB)	Right Bits Per Pixel (bpp)			
		Text	Vectors	Quadtree	Total
B16	37.22	2.13	0.02	0	2.15
QUAD	37.19	1.17	0.52	0.06	1.75
B16	31.07	0.80	0.02	0	0.82
QUAD	31.55	0.05	0.50	0.06	0.61
B16	24.82	0.12	0.02	0	0.14
QUAD	25.23	0.05	0.22	0.04	0.31

Table 1. Comparison of bits spent on coding fixed block size and quadtree representations of Train & Tunnel.

As can be seen from the table, for medium and high bit rates, the quadtree estimation results in a much higher proportion of the bit budget being spent on coding the disparity vectors than for fixed block size methods. However, this is more than matched by a drop in bits needed to code the prediction residual, and an overall improvement results. At low bit rates, the proportion of bits spent coding texture in fixed size coding is much less, as disparity vectors must be coded in a lossless fashion. The potential savings in using quadtree estimation are therefore negated. Indeed, the overhead required becomes significant. Figure 2 shows an example of the quadtree

decomposition obtained using this technique, with Figure 3 showing the resultant disparity maps of the two estimation techniques.

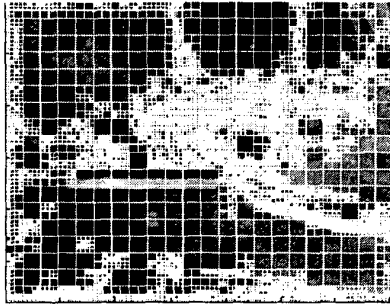


Figure 2 Quadtree Decomposition of Train & Tunnel right view, frame 0.

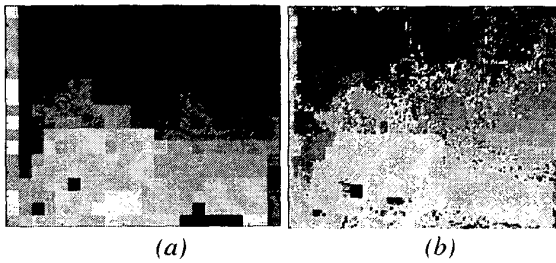


Figure 3 R-L Disparity Maps for (a) Fixed Block Size and (b) Quadtree Estimation of Train & Tunnel.

### 3 INTERMEDIATE VIEW SYNTHESIS

For accurate intermediate view synthesis, both the left-to-right and right-to-left disparity maps are required, such that areas only present in one image of the original stereopair (occluded areas) can be correctly synthesised at intermediate positions. One idea proposed in this paper is that only one disparity map needs to be reasonable detailed or sharp, and the other can just be an approximation coded at a much lower bitrate. Thus, a low resolution left-to-right disparity map was estimated at the encoder, again using quadtree estimation, and transmitted.

The intermediate view synthesis is then a relatively simple process which has been well documented [8,9,10], whereby the image intensities are projected onto the intermediate view plane at positions governed by the intersection of the disparity vectors with that plane. Results for the CLAUDE2 image set are shown in Figure 5. The decoded left image is shown in (a) at 31.93 dB PSNR. The decoded right image (b) is at 32.56dB PSNR. The synthesised intermediate view luminance PSNR was reasonable considering it is obtained for only a small overhead over

simple stereo image coding, at 27.86 dB, with a few artefacts visible around the edges of the foreground object. For this experiment, a value of 0.07 bits per pixel was chosen for the low resolution left-to-right map. This compared to 0.41 bpp for the right-to-left map.

A block diagram of the method used for intermediate view synthesis using quadtree based disparity maps is shown in Figure 4 below.

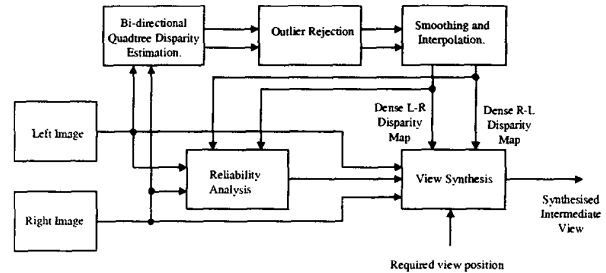


Figure 4 Intermediate View Synthesis Using Quadtree Disparity Maps.

The left-to-right and right-to-left disparity maps are produced using quadtree disparity estimation, with one vector per block. For intermediate view synthesis, dense maps containing one vector per pixel are required. To create these dense maps, firstly isolated outliers in the disparity maps are identified and removed. Then each pixel in the remaining blocks is set to the block disparity vector. Next, a median filter is applied to smooth out sharp, isolated discontinuities in the disparity field. This is done based on the assumption that all objects in the image are larger than a certain small size and a reasonable distance from the camera (i.e. not right next to the camera lens).

The disparity fields are interpolated using a simple bi-linear interpolation method, creating a pair of dense disparity maps. A reliability check is then carried out, based on bi-directional consistency in the disparity fields, and unreliable vectors identified. The reliability check is violated if

$$d_{i-1,j} - d_{i,j} > T, \quad (1)$$

where T is an integer threshold. T was chosen as 2 in this case.

Finally, the intermediate view synthesis is performed based on disparity controlled projection of the intensity values from the left and right images, using equation (2) below.

$$I(i, j) = (1 - \alpha).L(i, j + d_l) + \alpha.R(i, j + d_r) \quad (2)$$

where

$$\begin{aligned} d_l &= -\alpha \cdot d_{l-r} \\ d_r &= -(1-\alpha) \cdot d_{r-l} \end{aligned} \quad (3)$$

and  $d_{l-r}$  is the left to right disparity vector intercepting the intermediate image plane at point  $(i,j)$ . Similarly,  $d_{r-l}$  is the right to left disparity vector intercepting the intermediate image plane at point  $(i,j)$ . The intermediate view position between the left and right camera positions is represented by  $\alpha$ , where  $0 \leq \alpha \leq 1$ , and 0 corresponds to the left image, 1 the right.

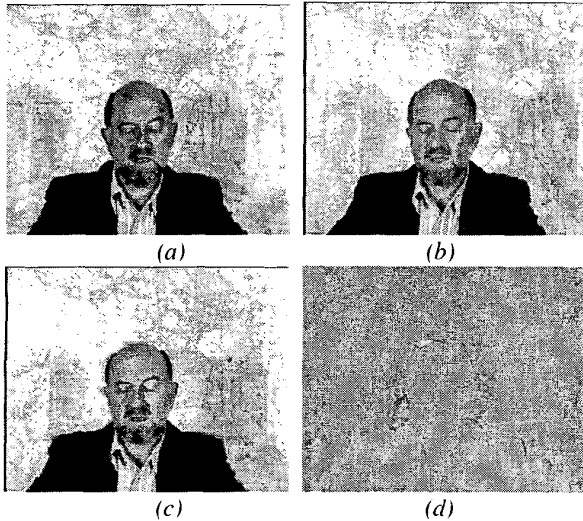


Figure 5 Synthesis Results for CLAUDE2: (a)Decoded left view @ 31.93dB PSNR, (b)Decoded right view @ 32.56dB PSNR, (c) Synthesised Intermediate View @ 27.86dB PSNR, (d) Difference between (c) and original intermediate view.

#### 4 CONCLUSIONS

The work described in this paper has shown that it is possible to use block based disparity maps to achieve reasonable quality intermediate view synthesis. It has also shown that it is possible to use a quadtree structure to efficiently represent the disparity information, and at the same time achieve efficient stereo image compression, with very low overhead required to perform synthesis of the intermediate views between the coded stereo pair. This method shows promise for efficient coding of multiview images, but more work needs to be undertaken in the areas of occlusion detection and outlier rejection. Further work will focus on improving the quality of the synthesised intermediate views, and in using the quadtree structure for possible scalable stereoscopic image compression strategies. The techniques will also be extended to include multiview video compression. Trade-offs between the

number of bits spent on coding the stereo image pair and that in coding side information required for good quality multiview synthesis should be investigated.

#### ACKNOWLEDGEMENTS

This work was supported by the Virtual Centre of Excellence in Digital Broadcasting & Multimedia Technology Ltd. The images used were supplied by CCETT, France.

#### REFERENCES

- [1]: J.-R. Ohm, K. Grüneberg, E. Hendriks, E. Izquierdo M., D. Kalivas, M. Karl, D. Papadimitos, A. Redert : "Realtime hardware system for stereoscopic videoconferencing with viewpoint adaptation", *Signal Proc.:Image Comm.*, vol.14, nos.1-2, pp.147-171, Nov.1998.
- [2]: D. Tzovaras, N. Grammalidis, M.G. Strintzis, "Disparity field and depth map coding for multiview 3D image generation," *Signal Proc. : Image Comm.*, vol.11, no.3, pp.205-230, Jan. 1998.
- [3]: S.Sethuraman, M.W.Siegel and A.G.Jordan, "A multiresolutional region based segmentation scheme for stereoscopic image compression," *Proc. SPIE vol. 2419, Digital Video Compression: Algorithms and Technologies*, 1995.
- [4]: M.G. Perkins, "Data Compression of Stereopairs," *IEEE Trans. Commun.* Vol. 40, no. 4, pp. 684-696, April 1992.
- [5]: H. Aydinoglu, M.H.Hayes III, "Performance Analysis of Stereo Coding Algorithms," *Proc. ICASSP'96*, vol. 4, pp.2191-2194, 1996.
- [6]: A. Puri, R.V. Kollaris, B.G. Haskell, "Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4," *Signal Proc.:Image Comm.*, vol.10, nos.1-3, pp.201-234, July 1997.
- [7]: G.J. Sullivan and R.L. Baker, "Efficient quadtree coding of images and video," *IEEE Trans. Image Proc.*, vol.3, no.4, pp. 327-331, May 1994.
- [8]: E.M. Izquierdo, "Stereo matching for enhanced telepresence in three-dimensional videoconferencing", *IEEE Trans. Circuits and Systems for Video Tech.*, vol. 7, no. 4, pp.629-643, Aug. 1997.
- [9]: E. Izquierdo, J.-R. Ohm, "Image-based rendering and 3D modeling: A complete framework", *Signal Proc : Image Comm*, 15 (2000) 817-858.
- [10]: S. Pollard, M. Pilu, S. Hayes, A. Lorusso, "View synthesis by trinocular edge matching and transfer", *Image and Vision Computing* 18 (2000) 749-757.